ABS Data Compressing Kickoff Meeting - Status, Future Work, and Roadmap

Allen Huang

CIMSS/SSEC, University of WI-Madison

ABSTARCT

ABS will produce unprecedented volume of information rich atmospheric and environmental remote sensing data. In this part 1 presentation we are reviewing the current progress/approach of data compression and noise estimate. The future work to achieve both ground-based and on-board ABS data compression needs are also conversed. A roadmap designed to cover the end-to-end study needs is outlined as well.

CIMSS

Outlines

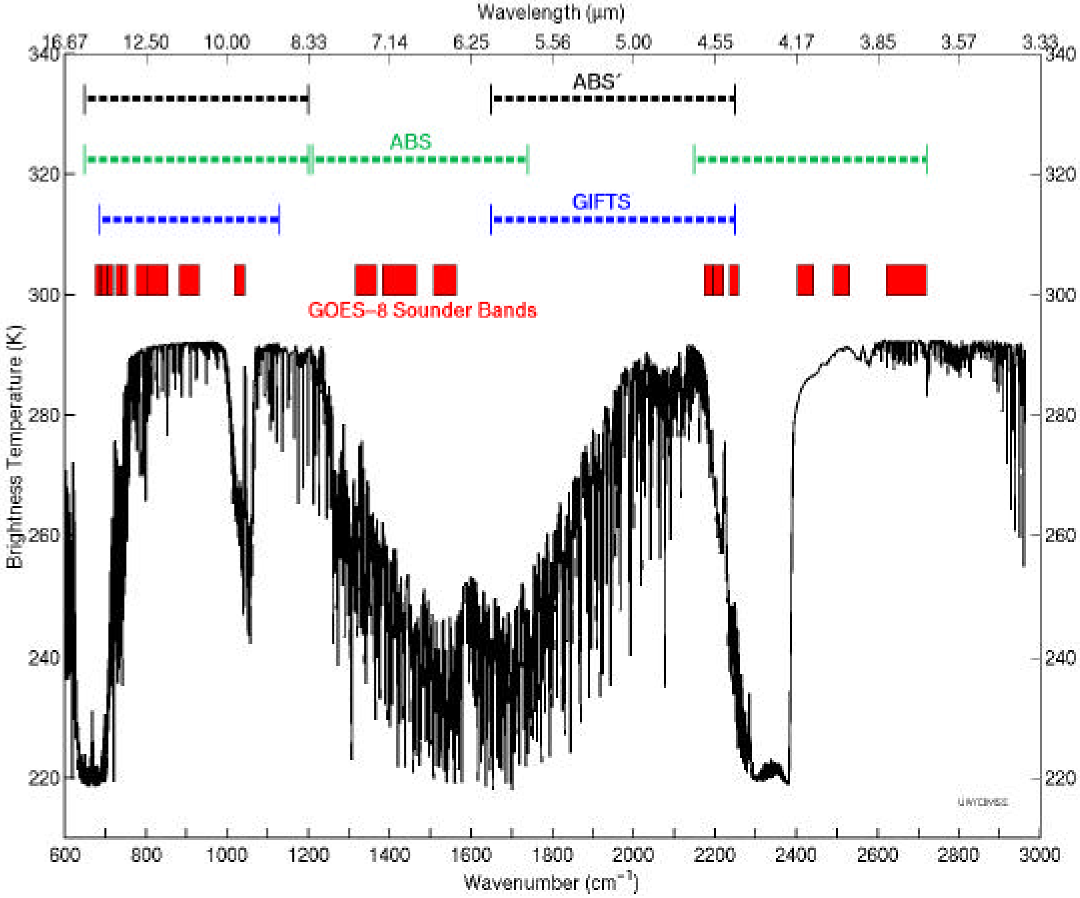
- Hyperspectral Data Information/Background
  - Data Volume; spectral, temporal, and gradient Inf. Con.
- GIFTS/NAST-I/AIRS data Compression Study Status
  - Will be discussed in details this afternoon
- Proposed Data Compression Approaches
  - Ground Based & On-board
  - Noise Estimation
- Roadmap for ABS Data compressing
  - Measurement Simulation, Data Compression Study, Defined Requirements, and Recommend Solution

CIMSS

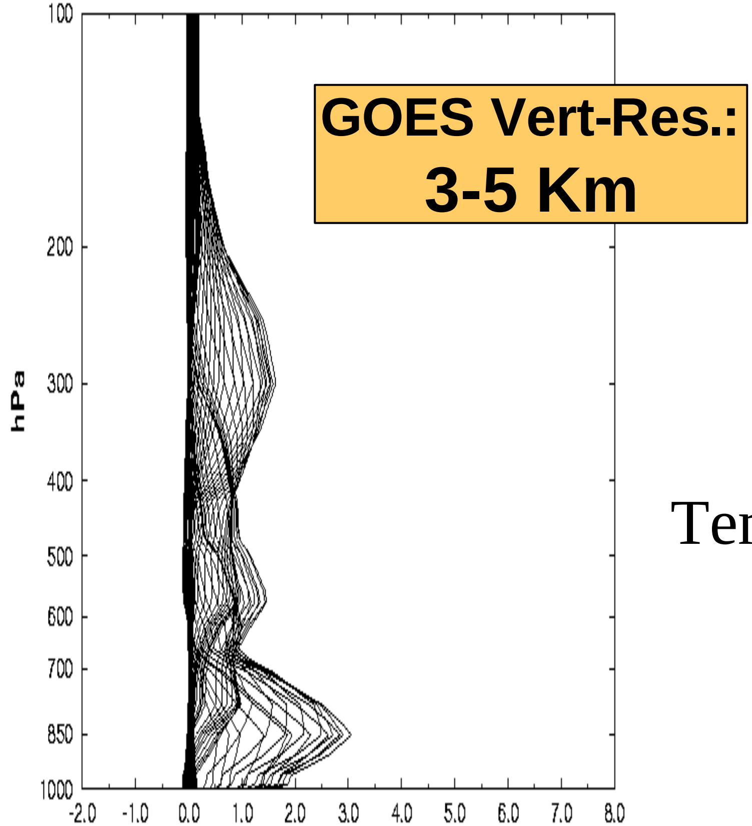
ABS Hyperspectral IR Data -

What to Expect -> Lots more data than we can handle

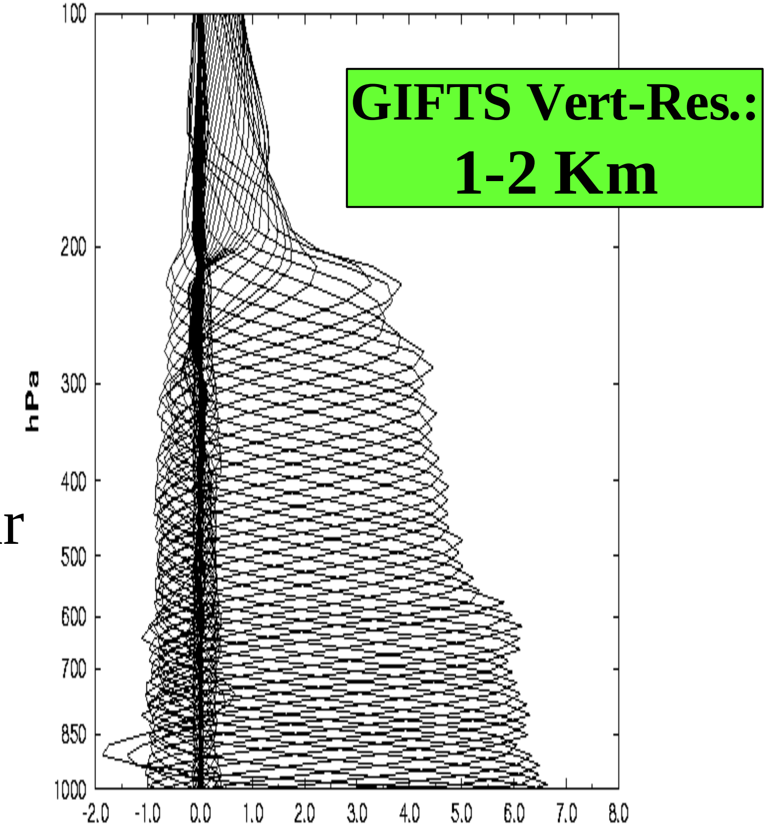
ABS/GIFTS ~ 3000 ch. per IFOV



Hyperspectral Data Information -
Spectral Information -> Vertical Resolution



Temperatur



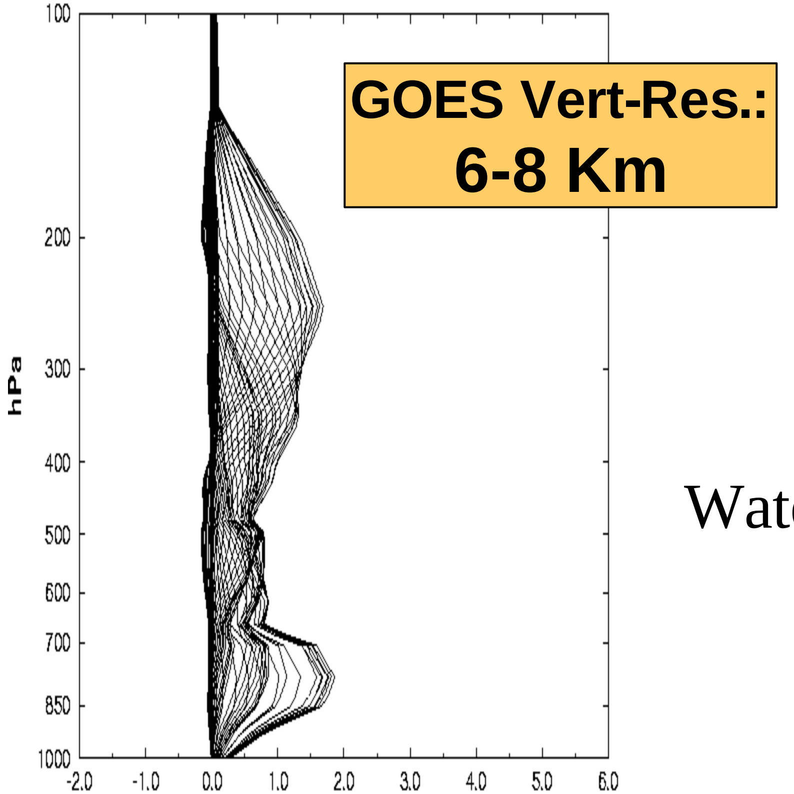
Current - GOES

~3 Pieces

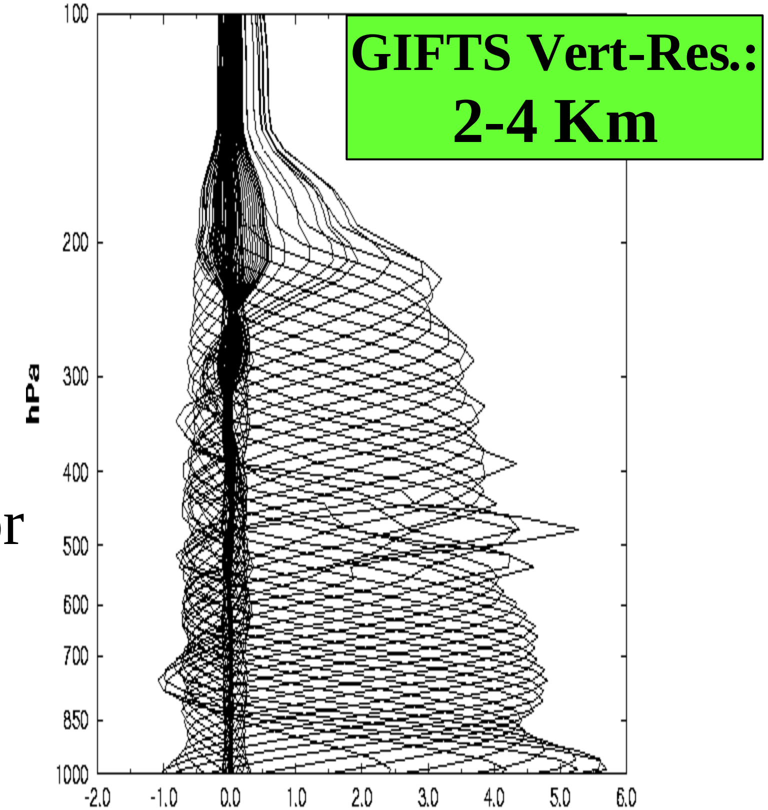
GIFTS

10-12 Pieces

Hyperspectral Data Information -
Spectral Information -> Vertical Resolution



Water Vapor



Current - GOES

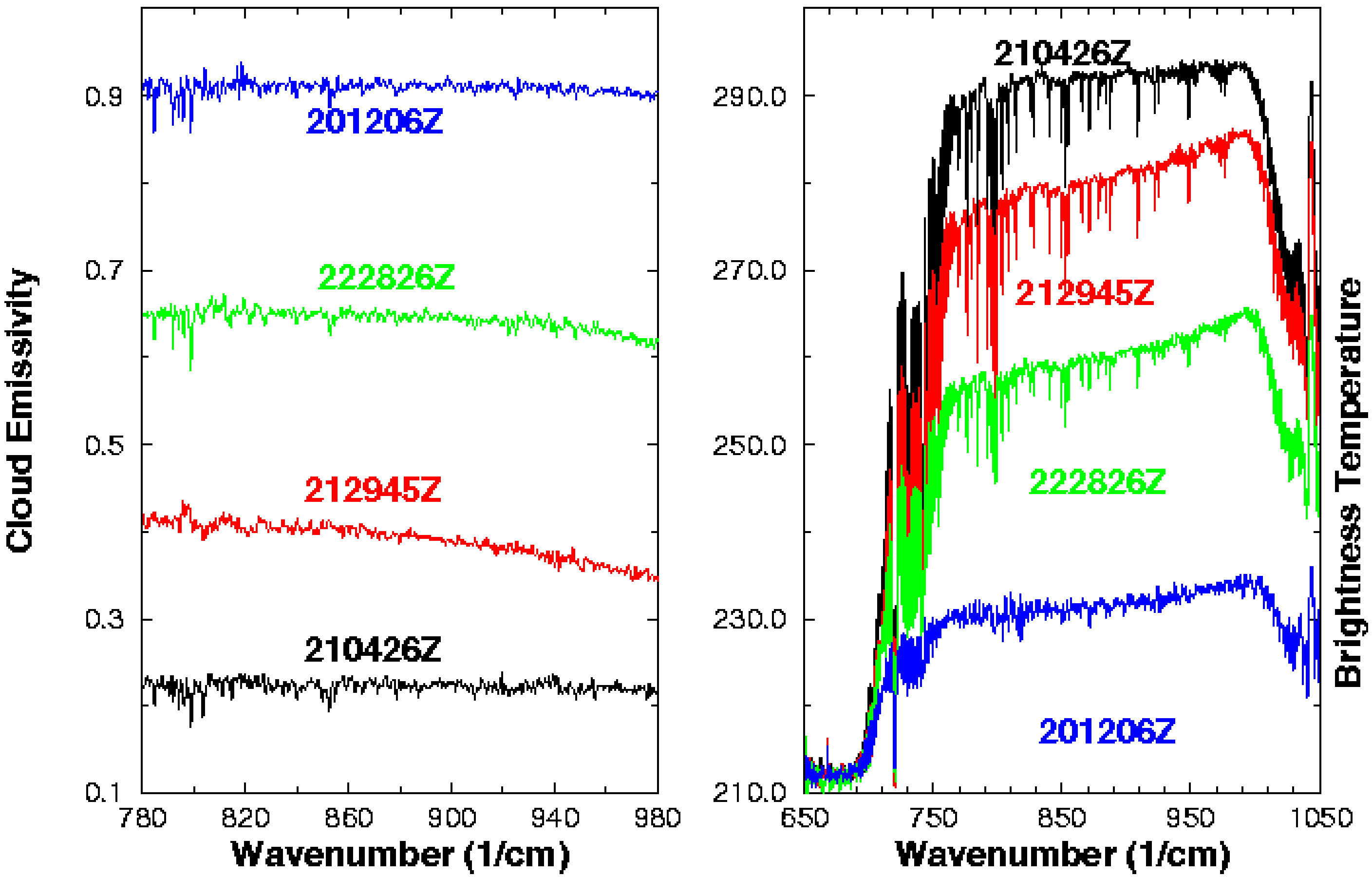
~2 Pieces

GIFTS

8-9 Pieces

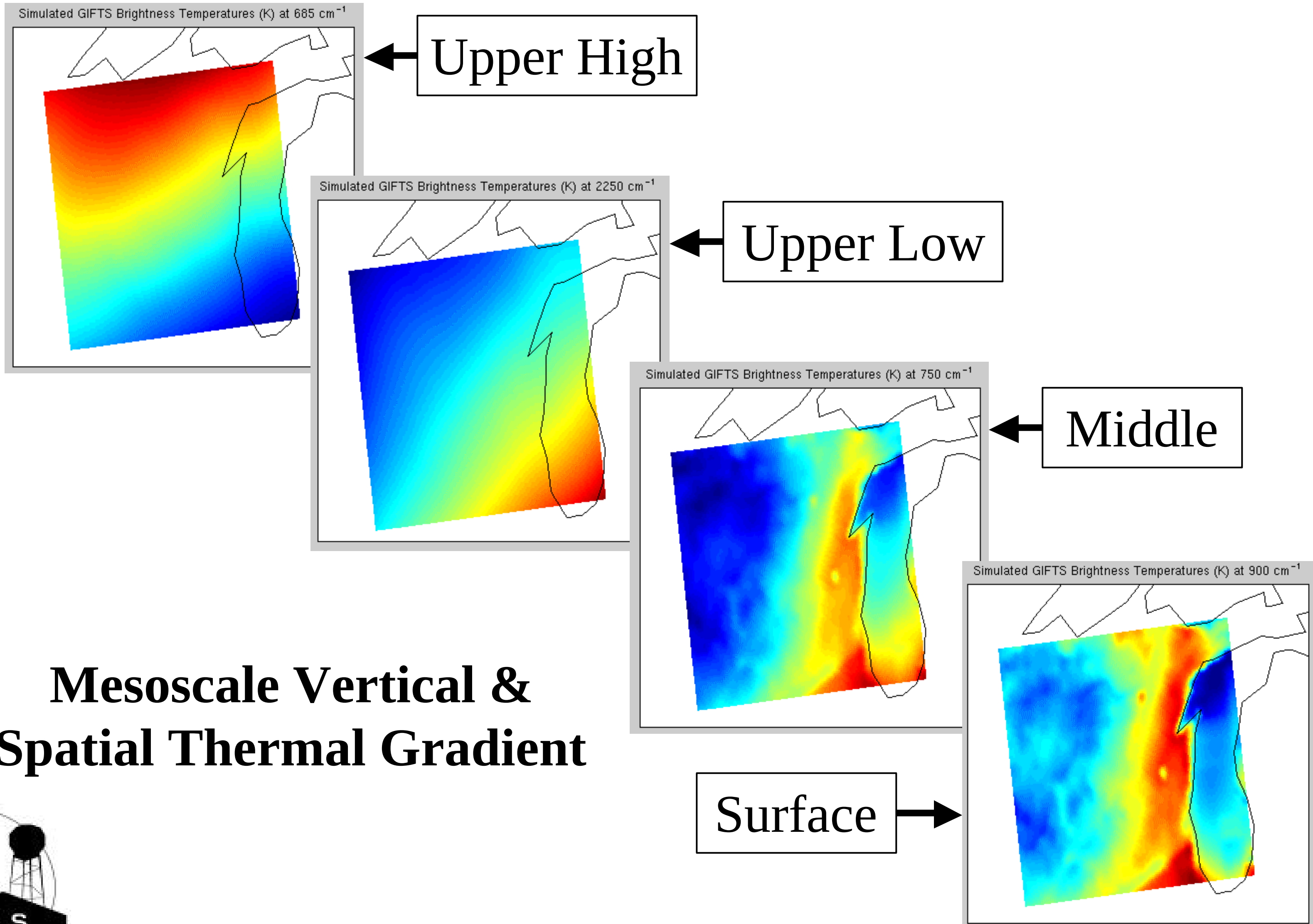
CIMSS

Hyperspectral Data Information - Spectral Information -> Marco & Micro Cloud Property



Hyperspectral Data Information -

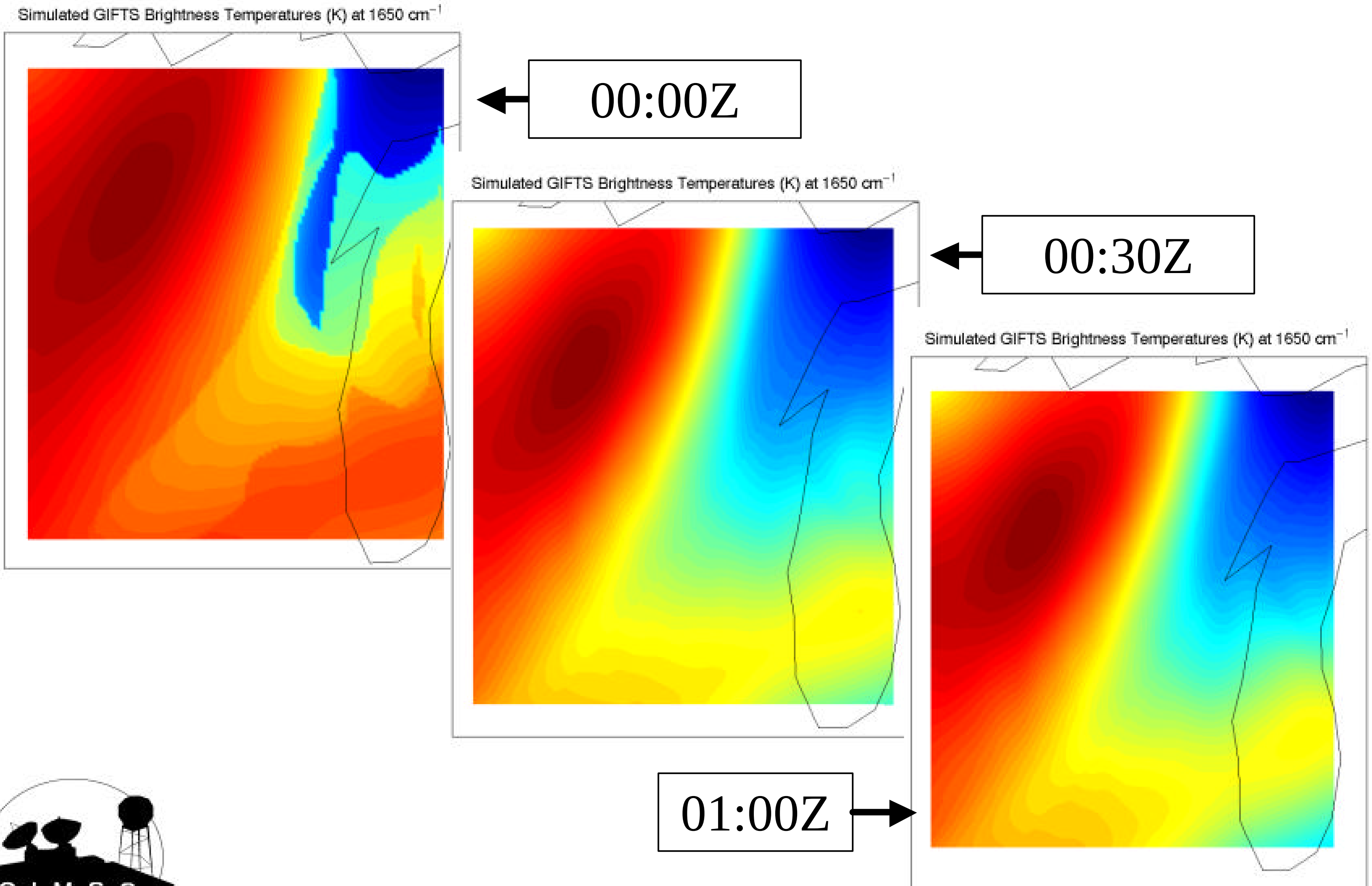
Spatial Information -> Gradient



Mesoscale Vertical &
Spatial Thermal Gradient

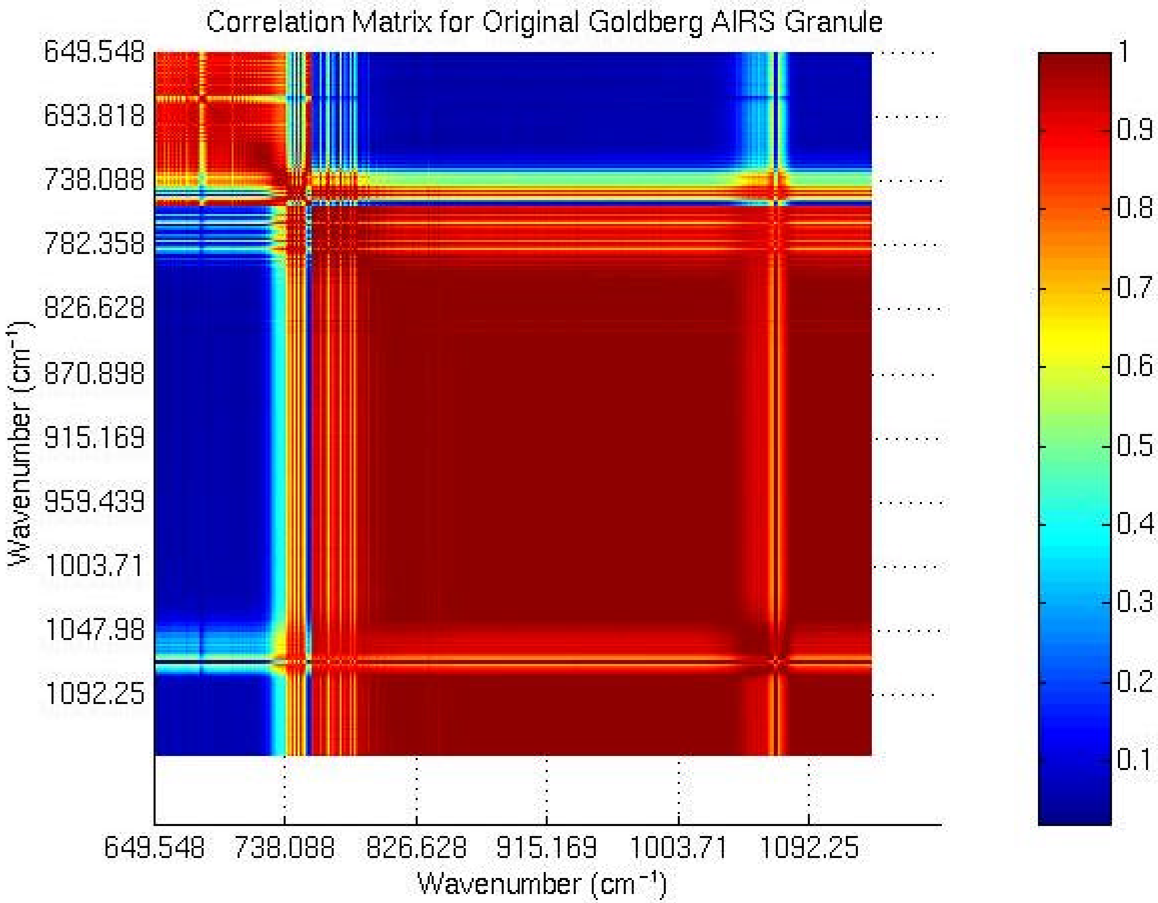
CIMSS

Hyperspectral Data Information -
Temporal Information -> Moisture Transport



CIMSS

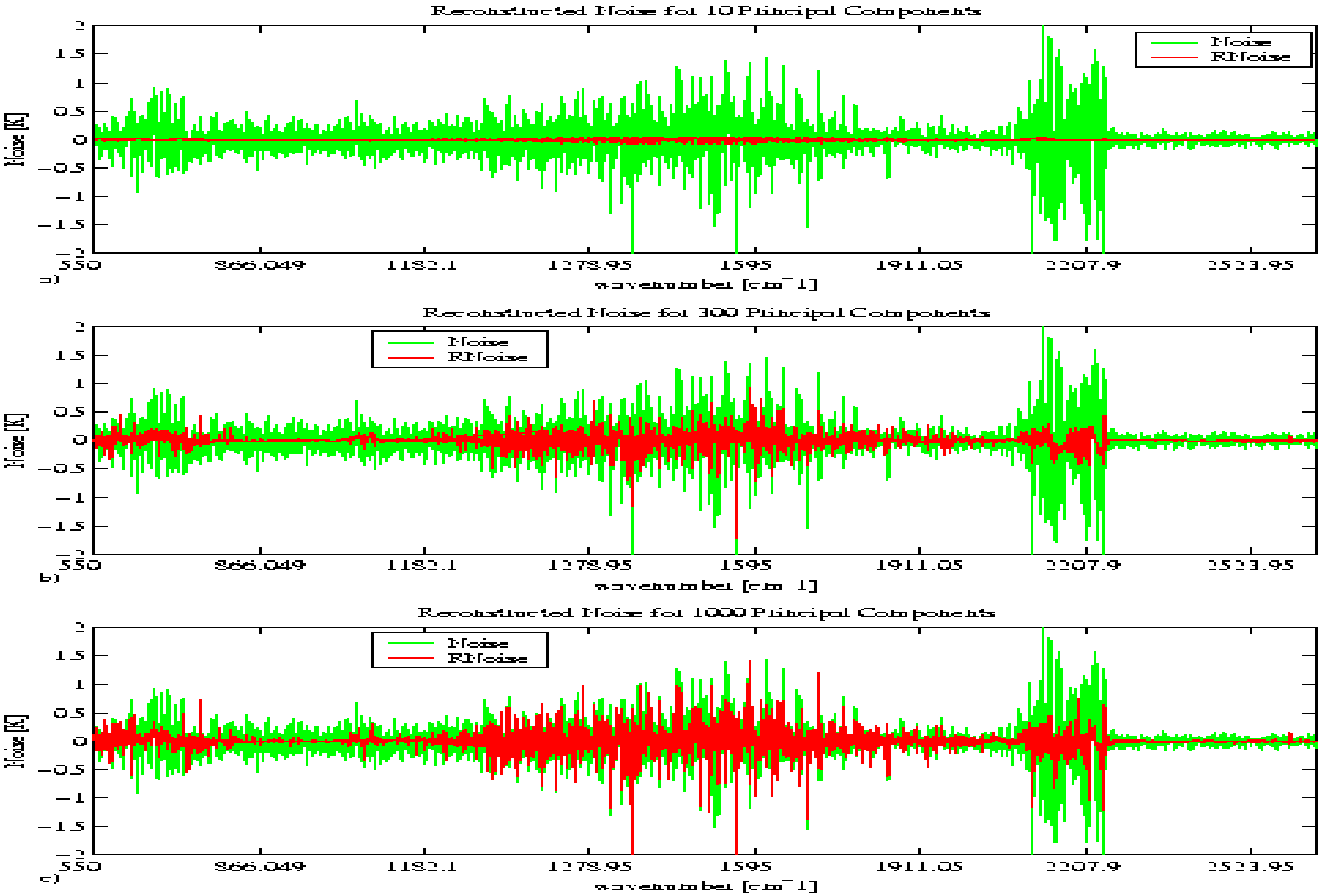
Hyperspectral Data Information - Spectral Correlation -> Data Redundancy?



What is PCA?

- PCA is Principal Component Analysis, a classical approach to the problem of linear (independent) feature extraction
- PCA essentially performs Singular Value Decomposition of the Covariance Matrix
- For Gaussianly distributed input, PCA extracts statistically independent features

Why Principal Component Compression (PCC) - Reason I : Noise Reduction



Noise (K)

10 PC

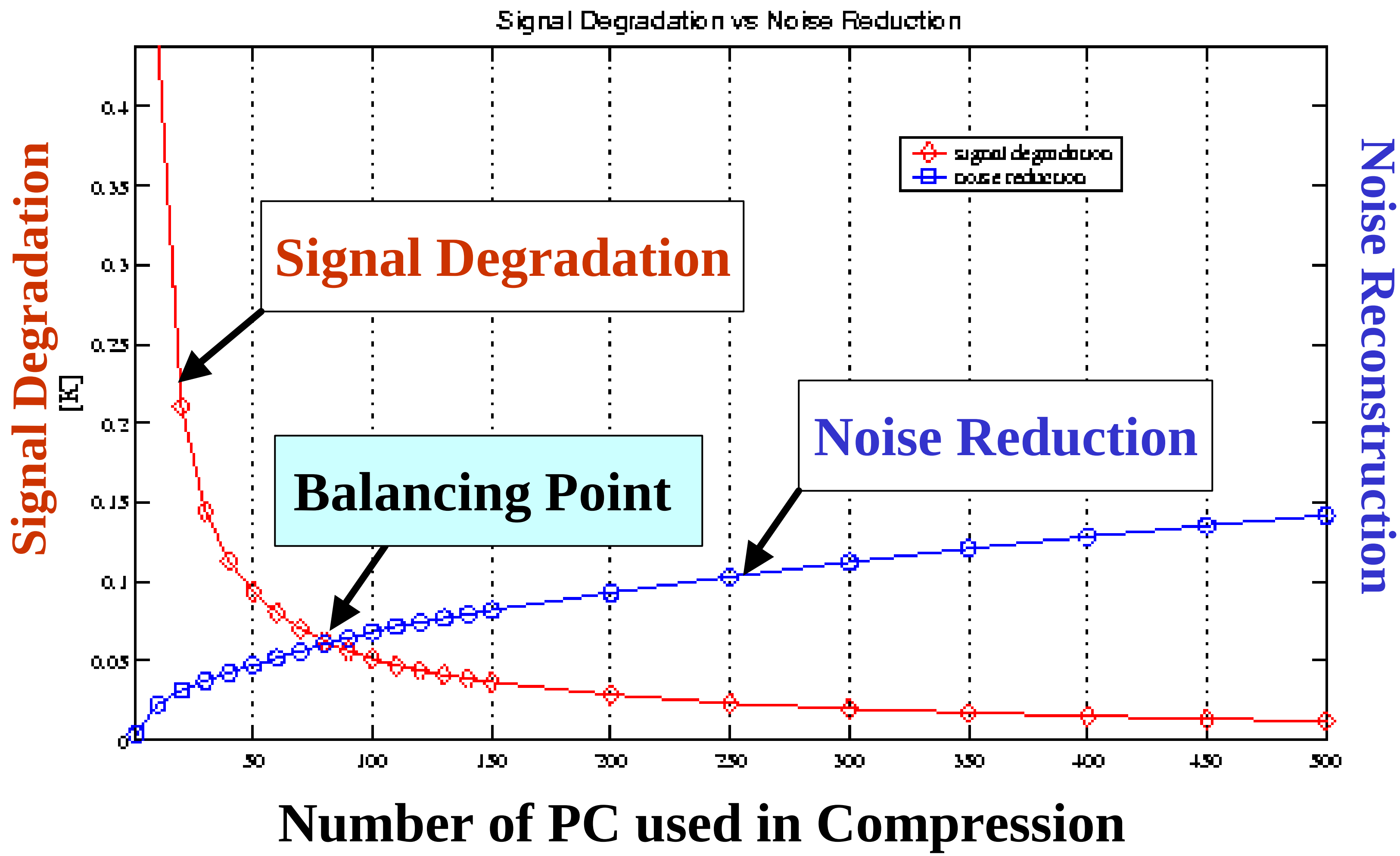
100 PC

1000 PC

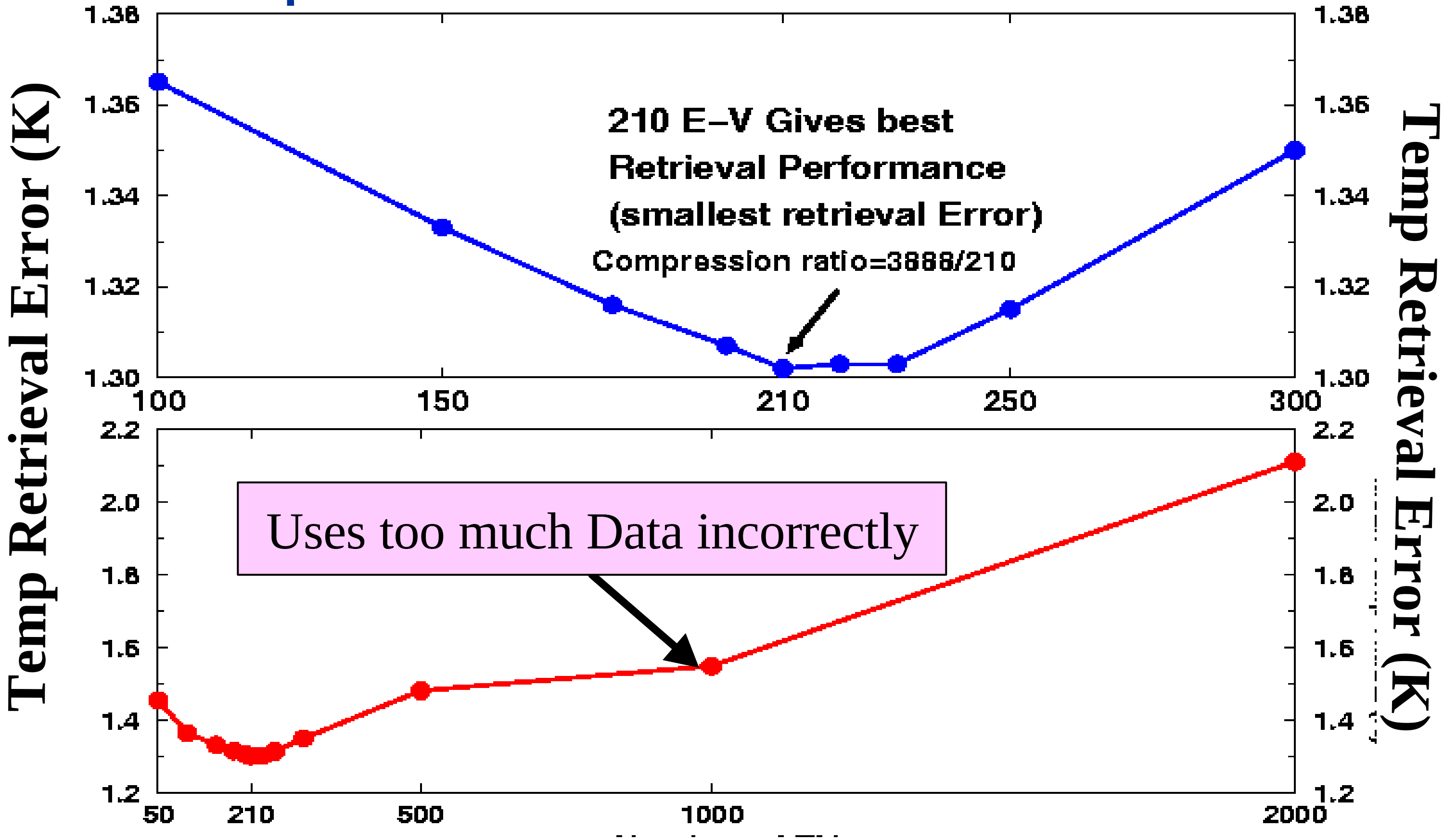
Wavenumber (1/cm)

Green-Single Noise Spectrum **Red-PCC Noise Spectrum**

Why PCC - Reason II : Tunable Signal Loss



Why PCC - Reason III : Discriminate Independent from Redundant Data



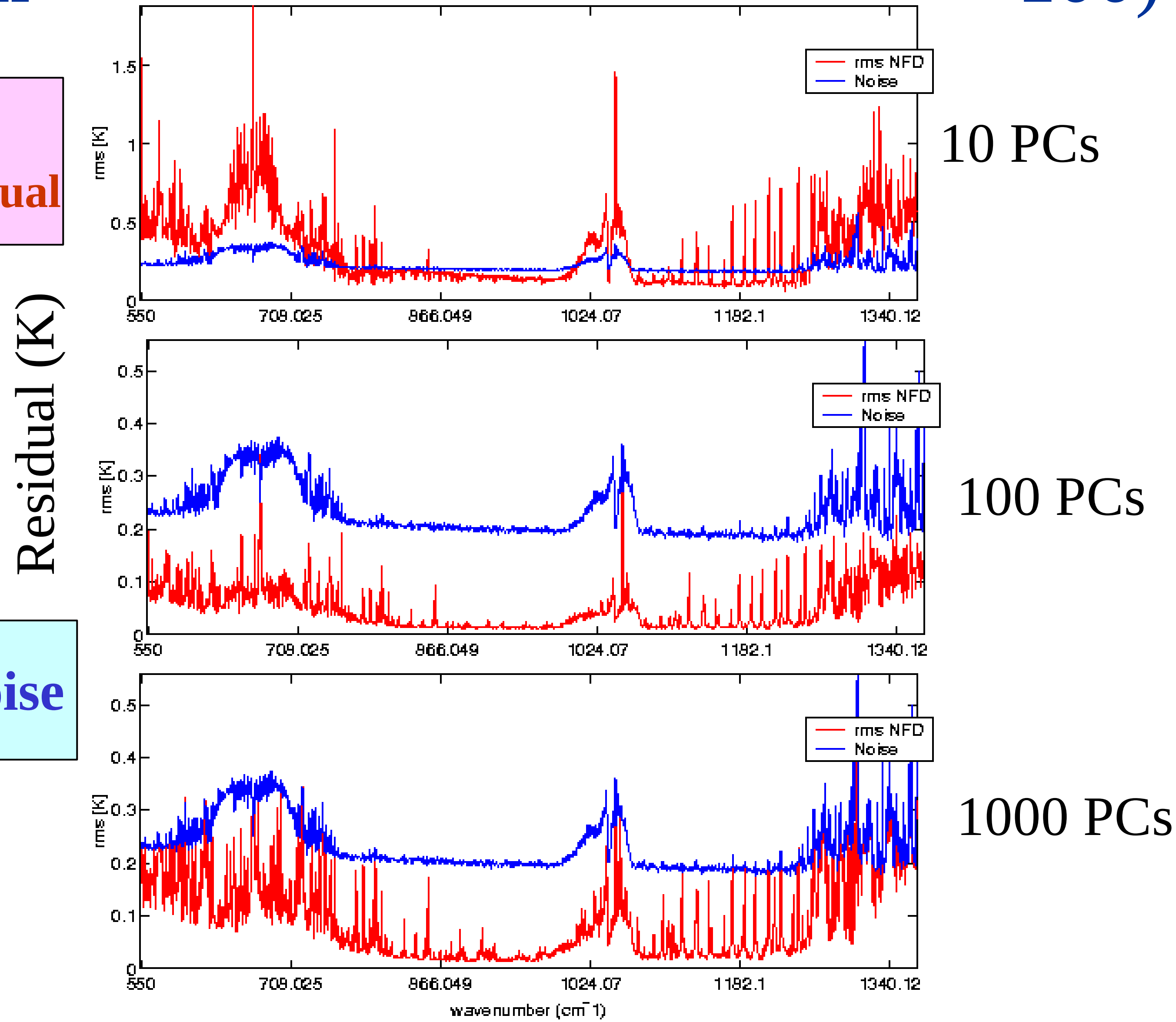
# of PC used

Uses of independent information only will improve retrieval performance

Why PCC - Reason IV : Consolidate Measurement Information (~2000 to ~100)

Blue - Noise
Red - PCC Residual

Residual < Noise



Principal Component Compression (PCC)

Pros & Cons

Cons:

1. **Requires PCA (Significant Computational Costs)**
2. **Requires Measurement Reconstruction**
3. **Representative Set Selection for PCs is an issue**

Pros:

1. **Noise Reduction/Estimation**
2. **Tunable Signal Loss**
3. **Discriminate Independent from Redundant Data**
4. **Increase Information Density**
5. **Intermediate Retrieval Processing step**
6. **Suitability to Available Computing Systems**

Hyperspectral Data Processing -

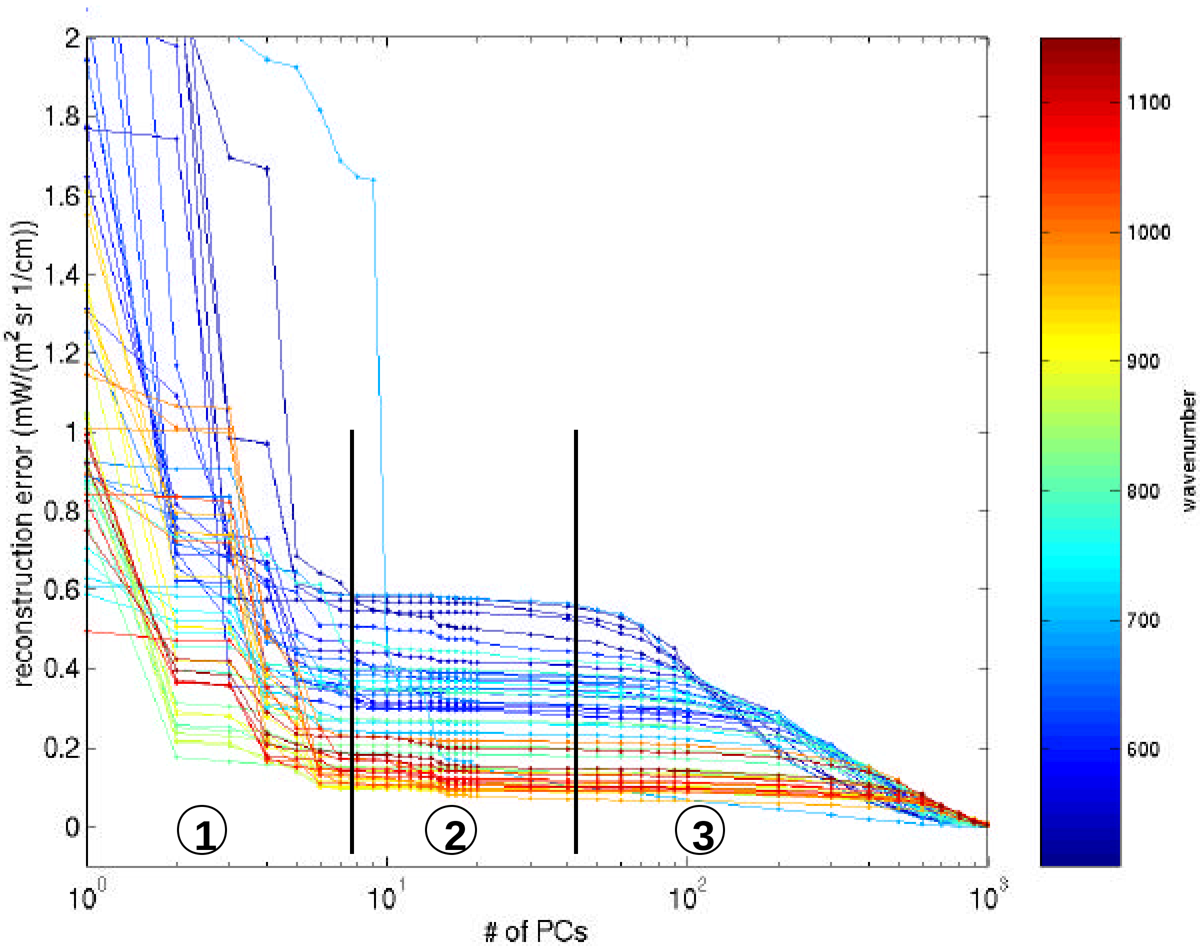
Why can we Estimate Noise Effectively ->

Signal & Noise are Estimated Separately

1 - Signal under represented

2 - Signal well represented; Noise filtered

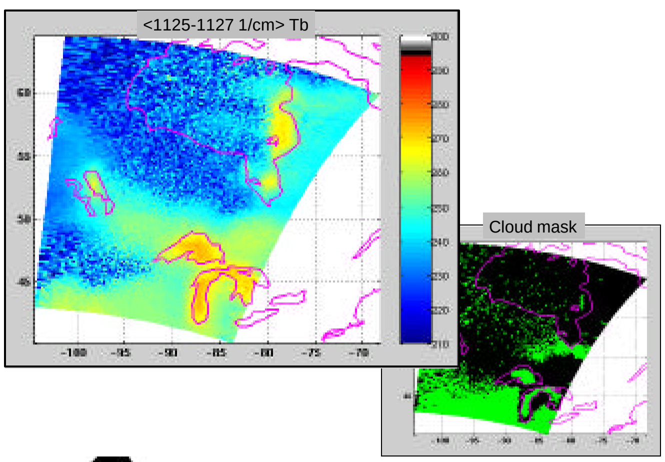
3 - Signal & noise both duplicated



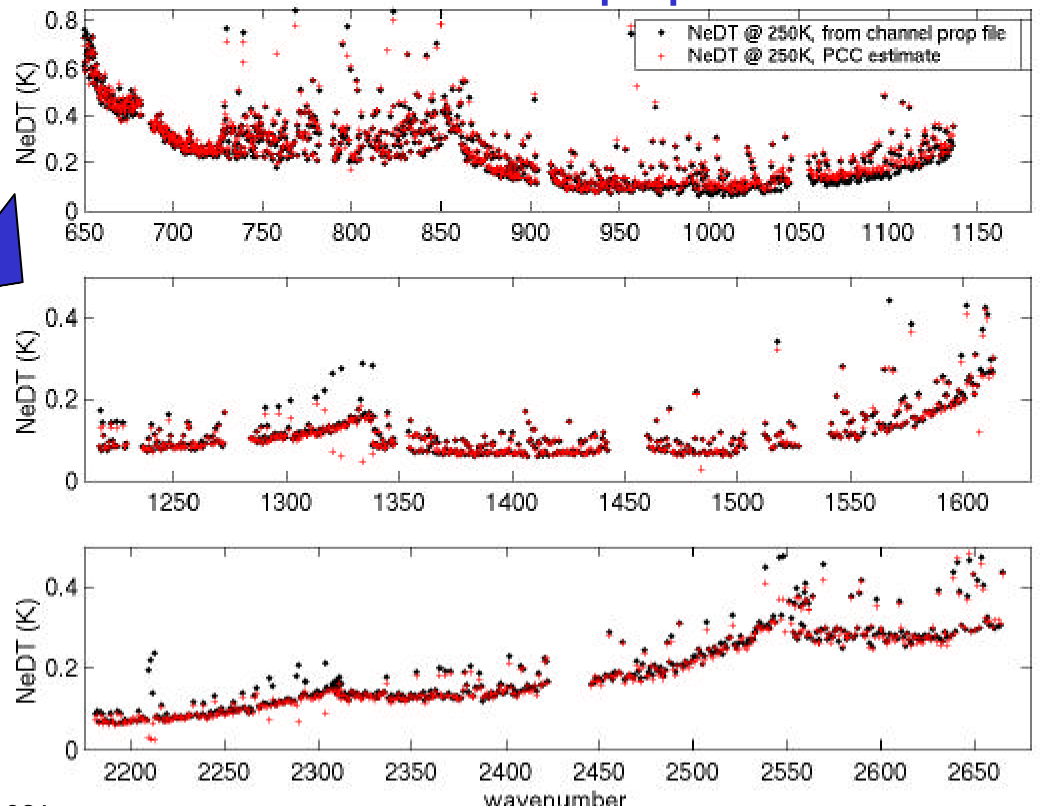
Hyperspectral Data Processing -
Why can we Estimate Noise Effectively ->
Noise are Well Estimated

Noise estimated well represent actual measurement noise

15-Dec-2000 granule 081



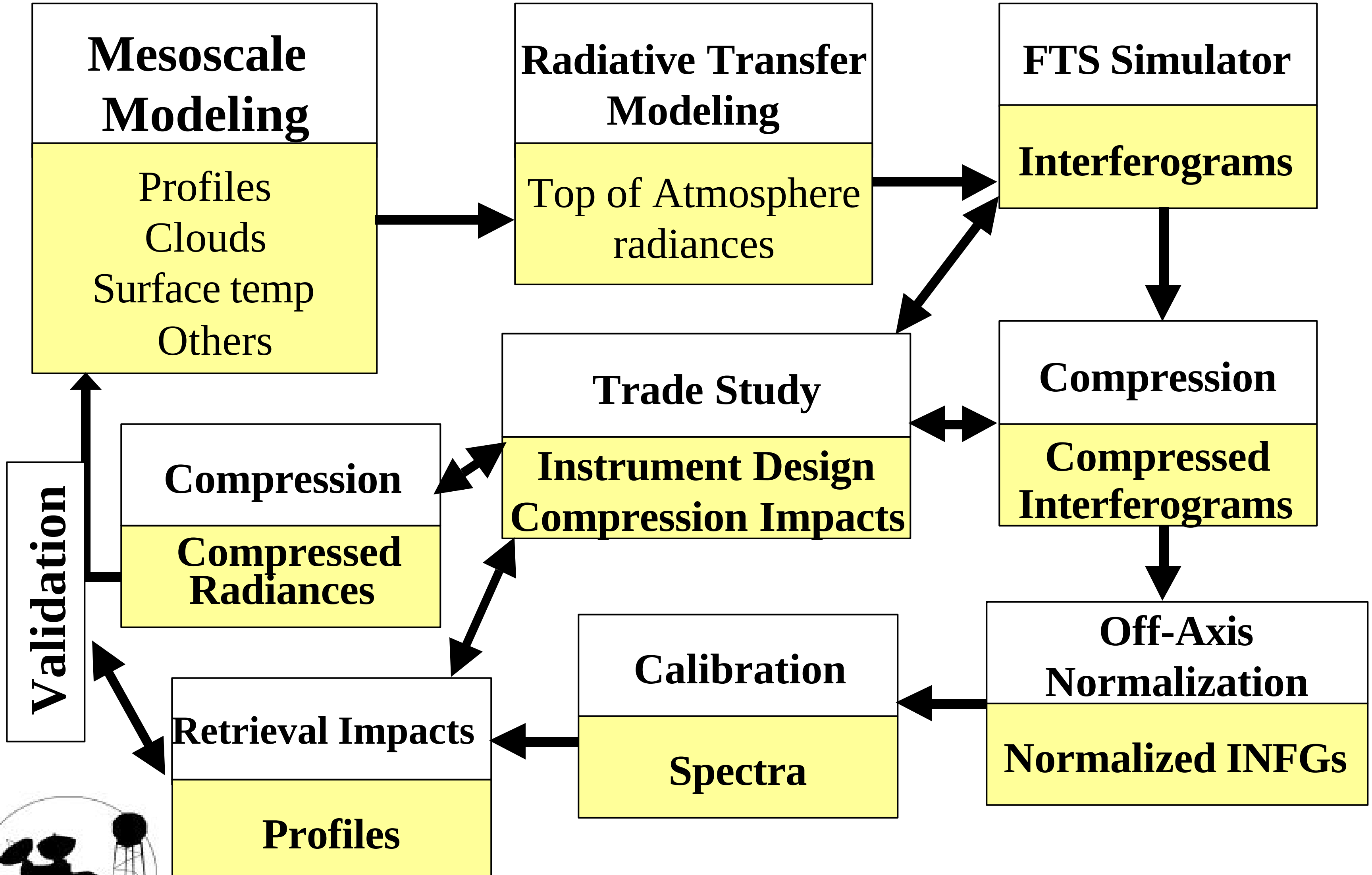
NeDT@250K: estimated from granule 081 and values from channel properties file



081

badFlag=0, radQual

ABS Measurement Simulation and Data Compression - Top Level Flow Diagram



ABS FTS Simulation

Input Calculated Top of Atmosphere Spectra

↓

- Add Instrument Background
- Apply Instrument Responsivity
- Apply Numerical Filter
- Apply Instrument Line Shape Function
- Apply Off-axis Interferogram Sampling

↓

Output ABS Interferograms or Spectra

CIMSS

ABS Measurement Noise Simulation

Input ABS ???

↓

- Apply ??
- Apply ??
- Apply ??
- Apply ??
- Apply ??

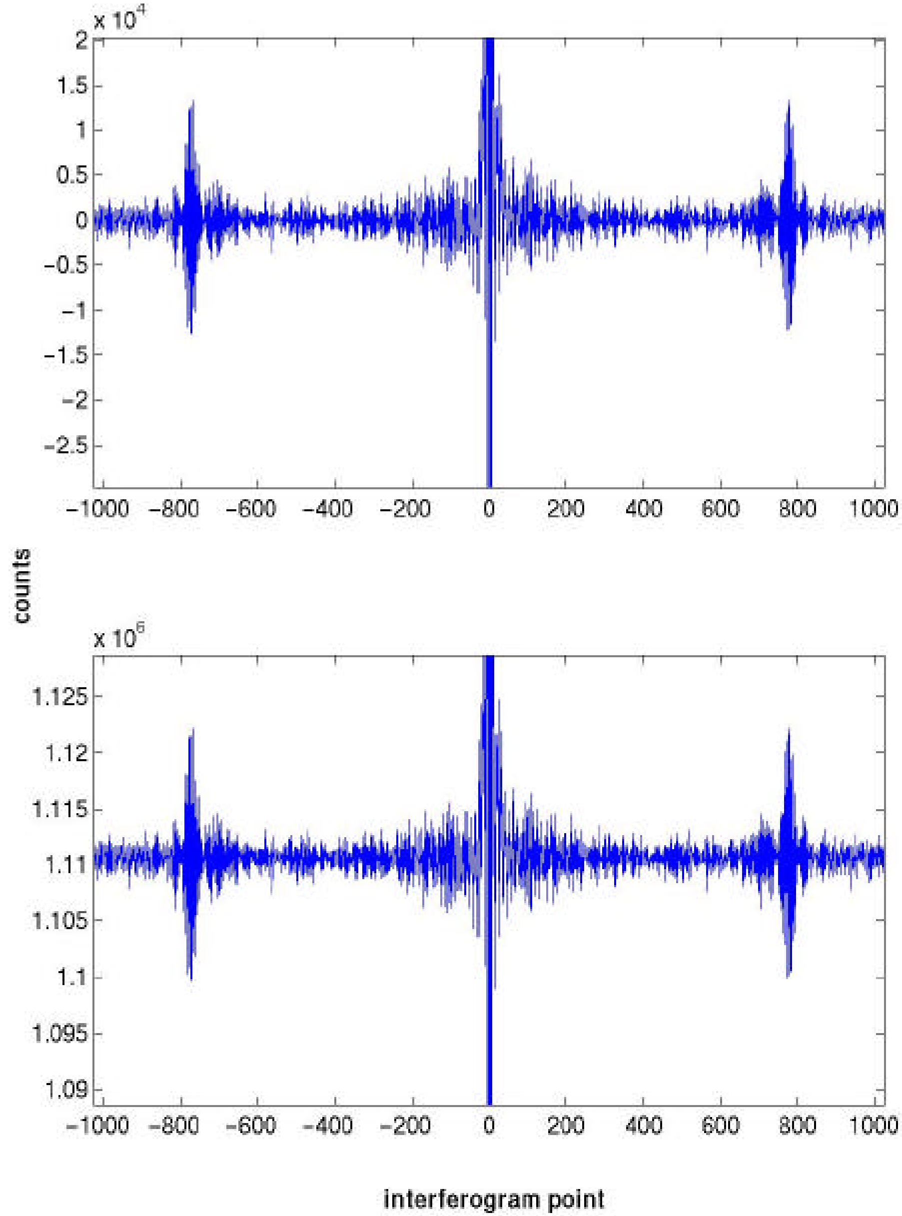
↓

Output ABS Noise Spectra

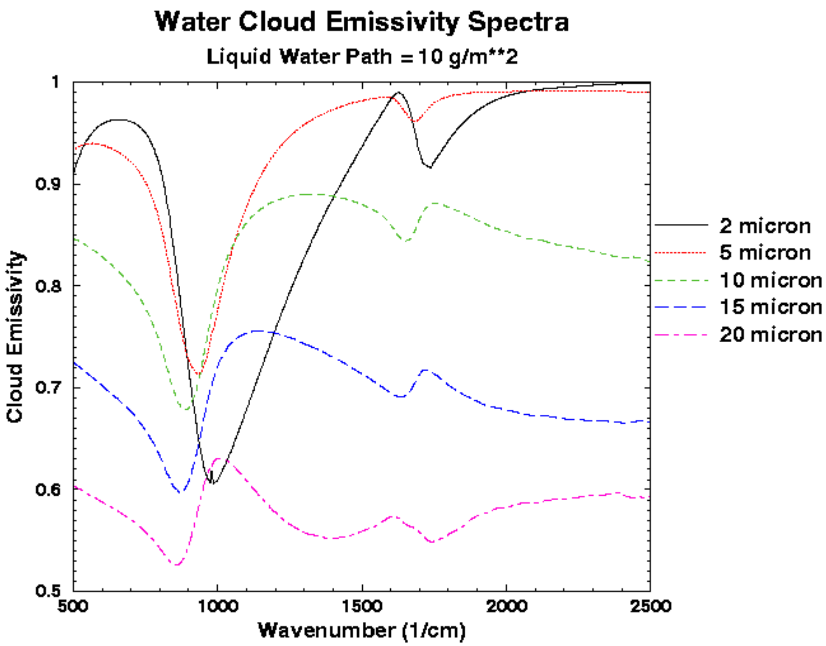
Example LW Interferogram with variable Gain and Offset

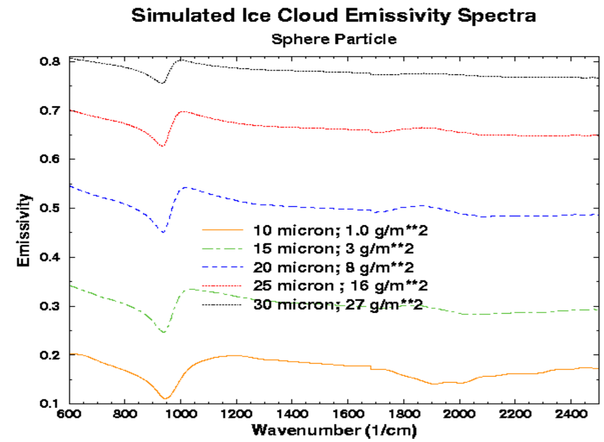
without effects

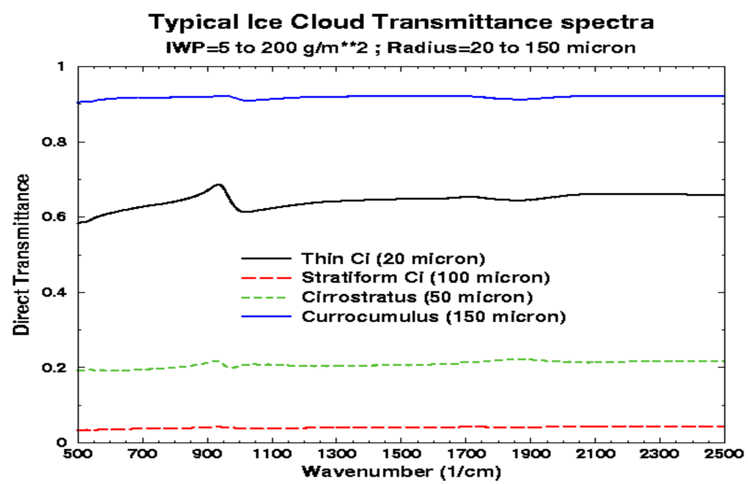
with effects



Parameterization of Ice and Liquid Cloud

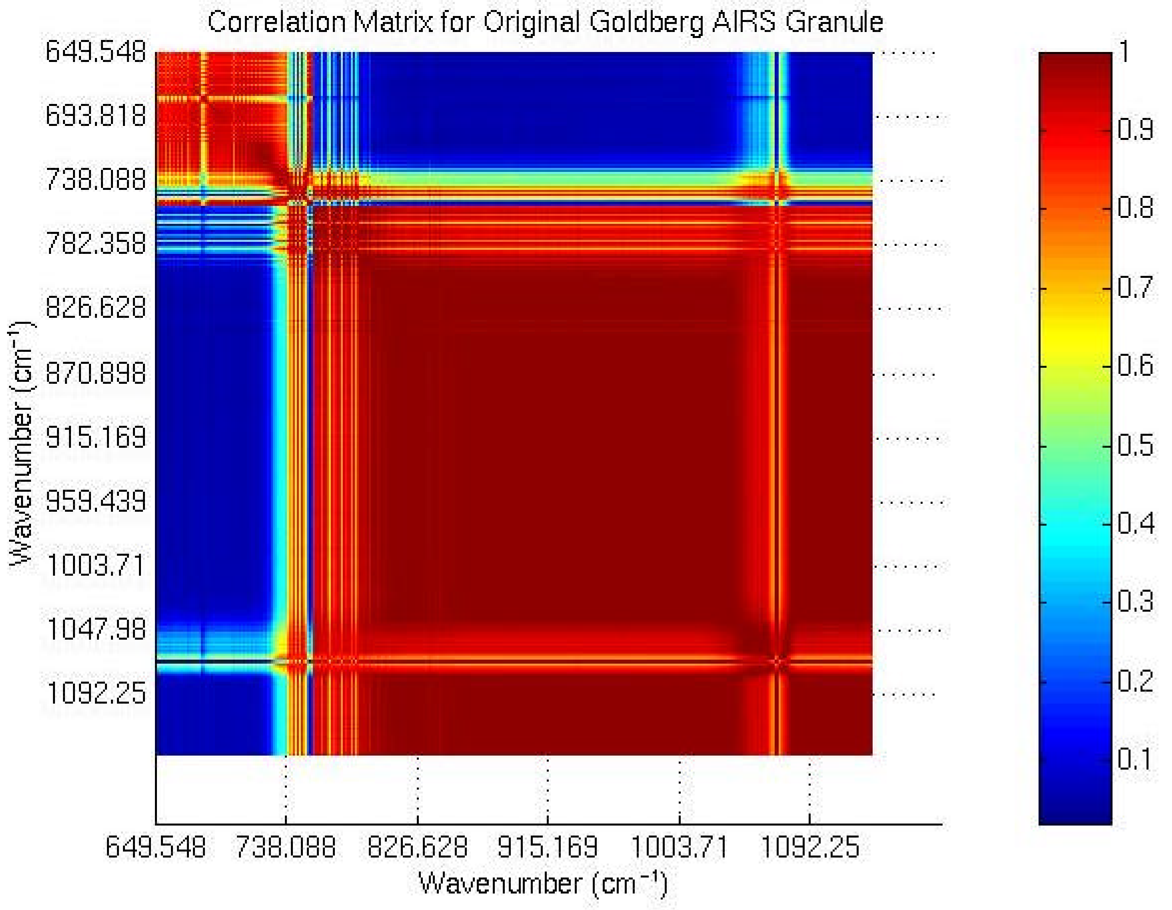






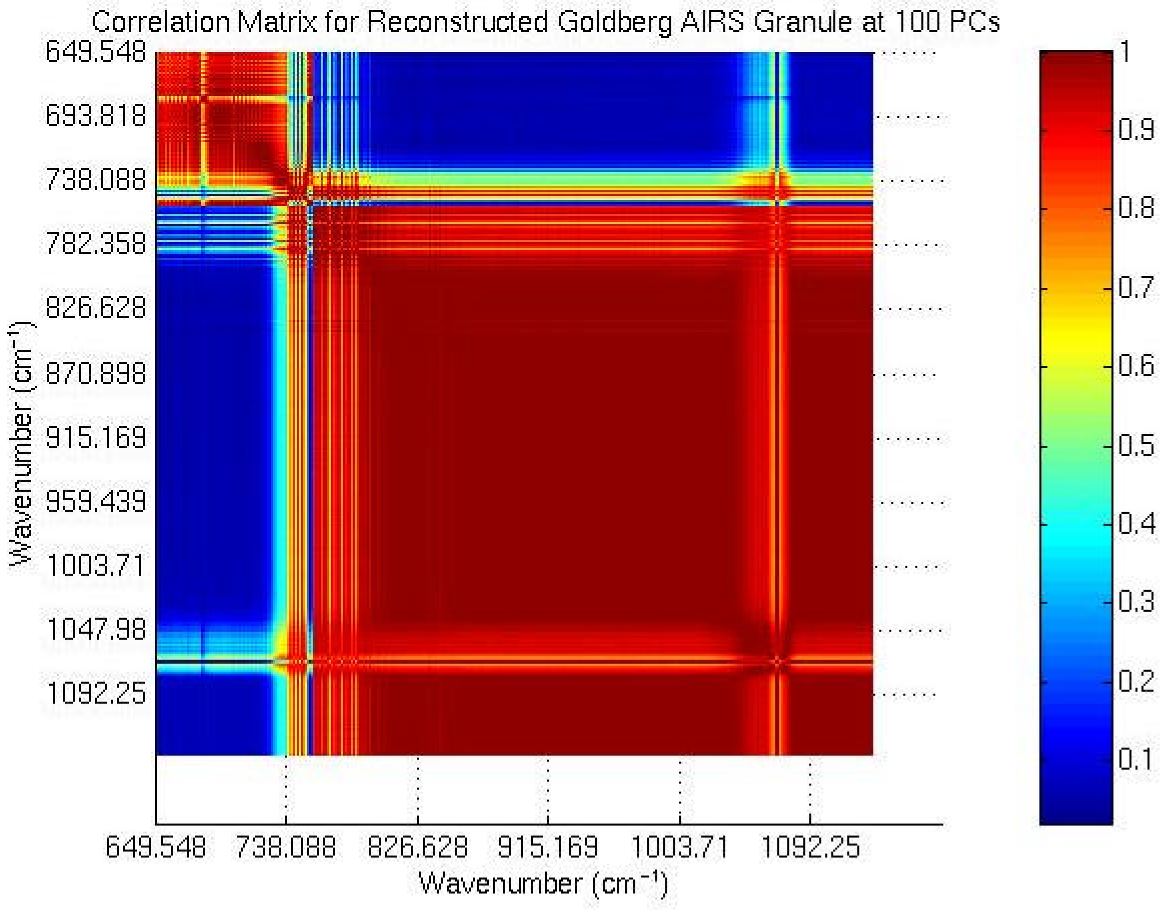
CIMSS

Raw Data Spectral Correlation

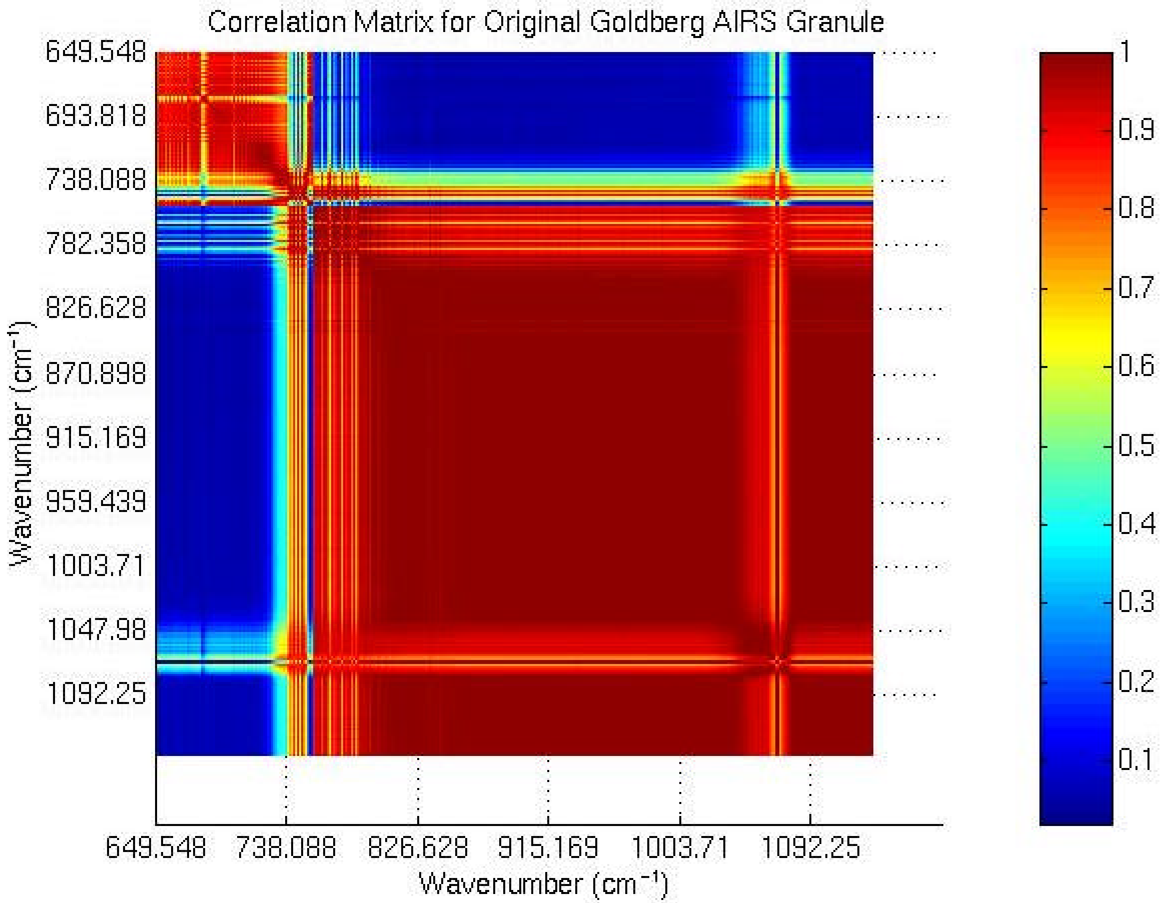


CIMSS

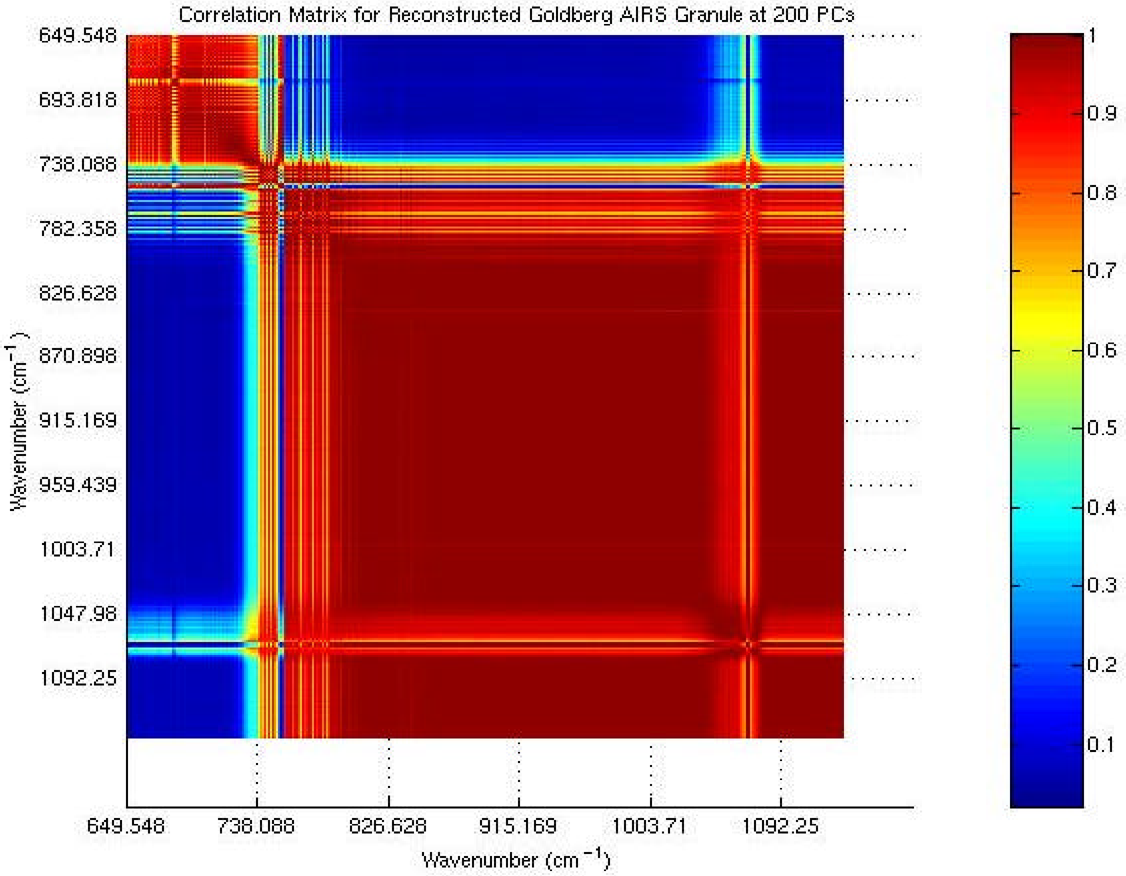
100 PC Compressed Data Spectral Correlation



Raw Data Spectral Correlation

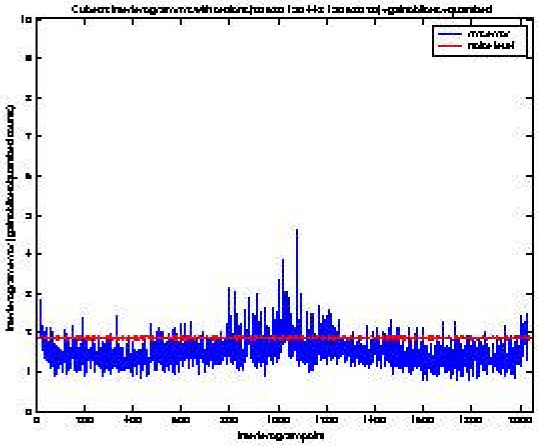


200 PC Compressed Data Spectral Correlation

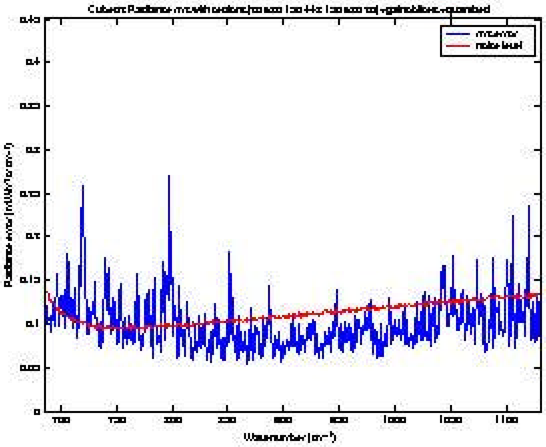


Compression Residuals - Raw or Calibrated Domains

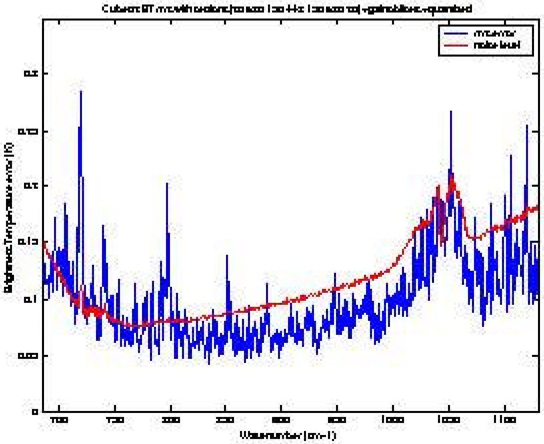
Interferogram



Radiance



Brightness Temp.



Definitions

Old DPC: Old Dependent Principal Component (fixed point segments)

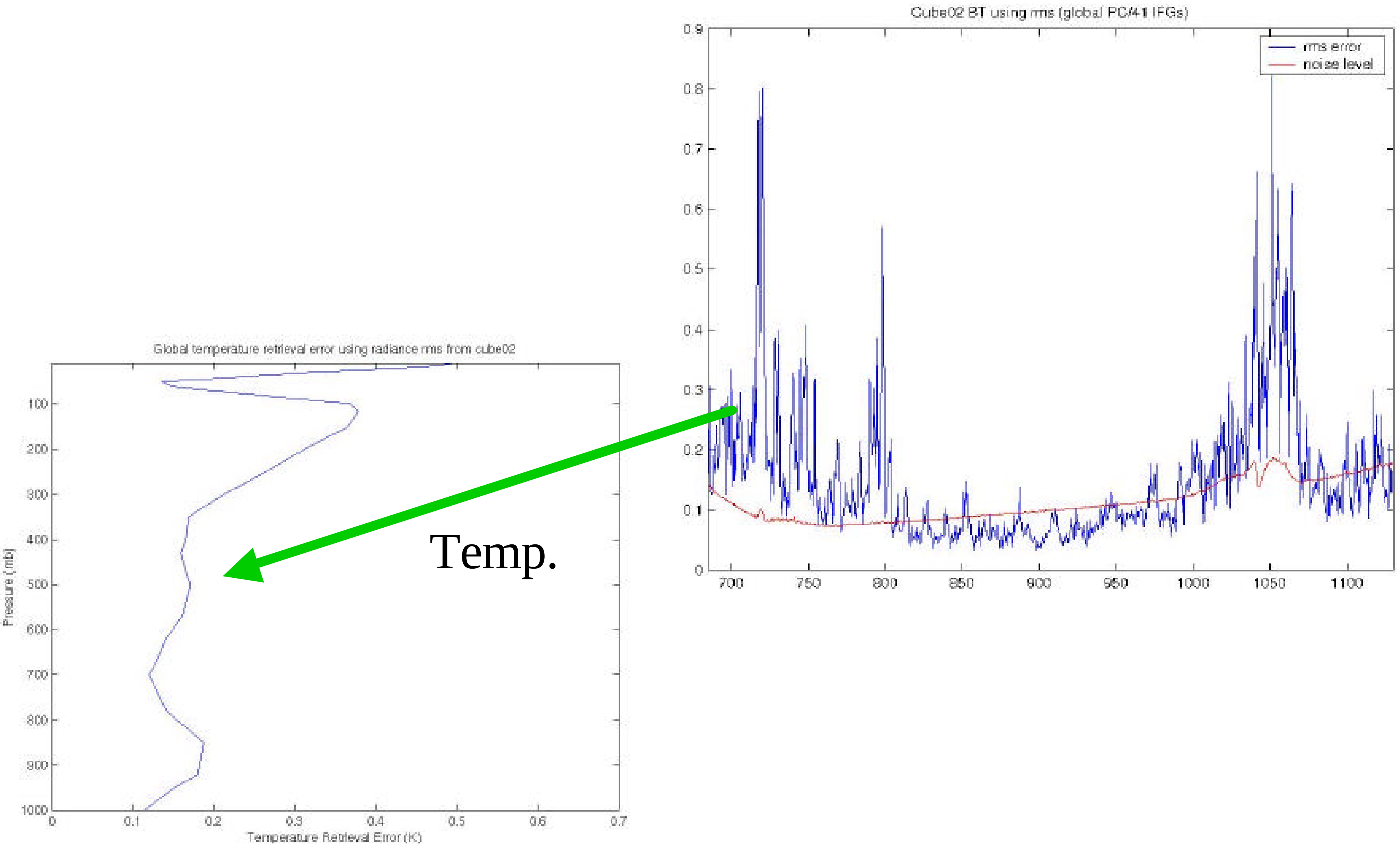
New DPC: New Dependent Principal Component (Variable point segments)

New HPC: New Hybrid Dependent & Independent Principal Component (Variable point segments and PCs are derived from on-line dependent and off-line independent historical/pre-computed data)

Note: 1. New DPC achieves better compression than old DPC
2. New HPC requires much less on-board processing than New DPC but degrades compression performance

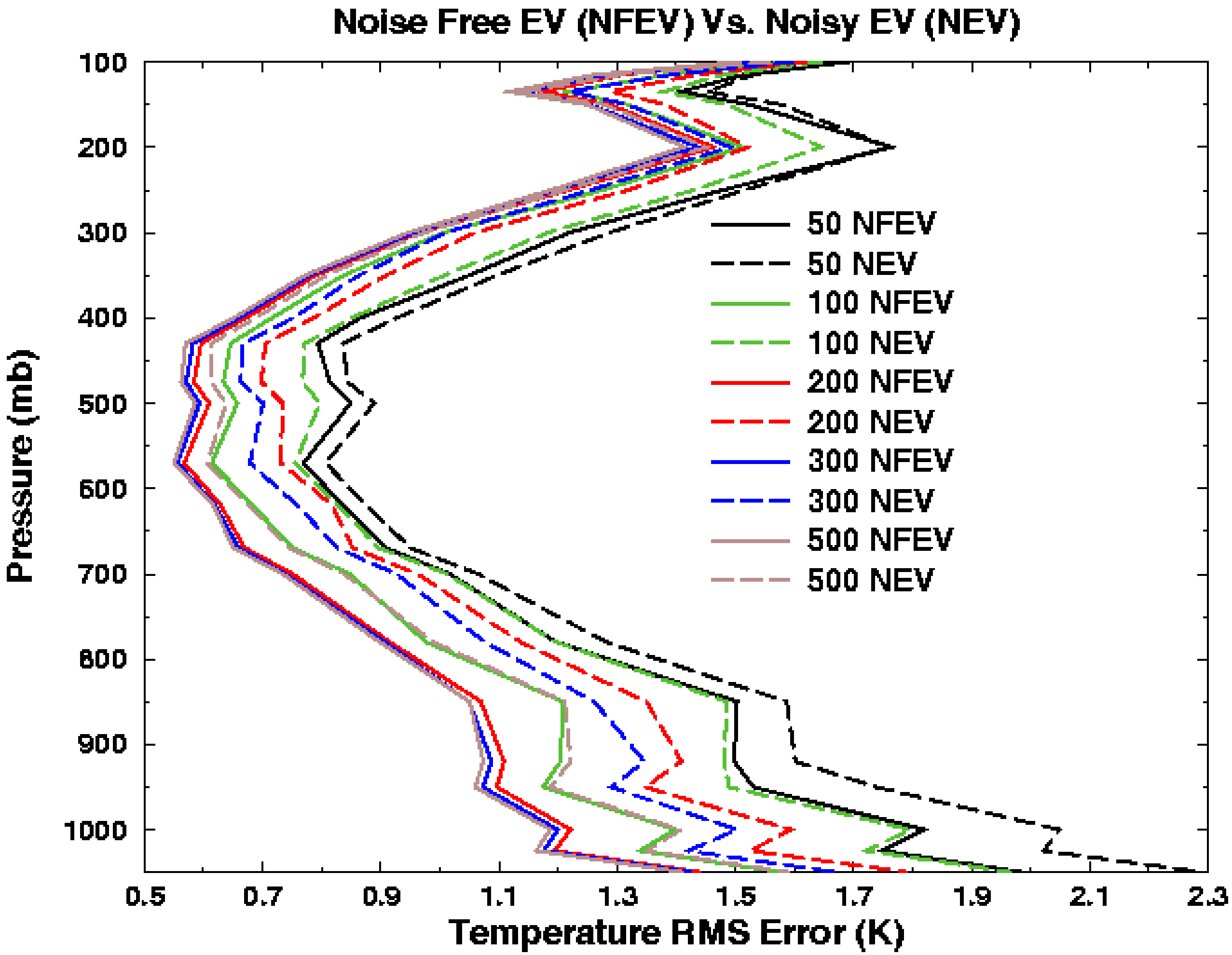
C I M S S

HPC Compression Impact on Retrieval - Cube 2



CIMSS

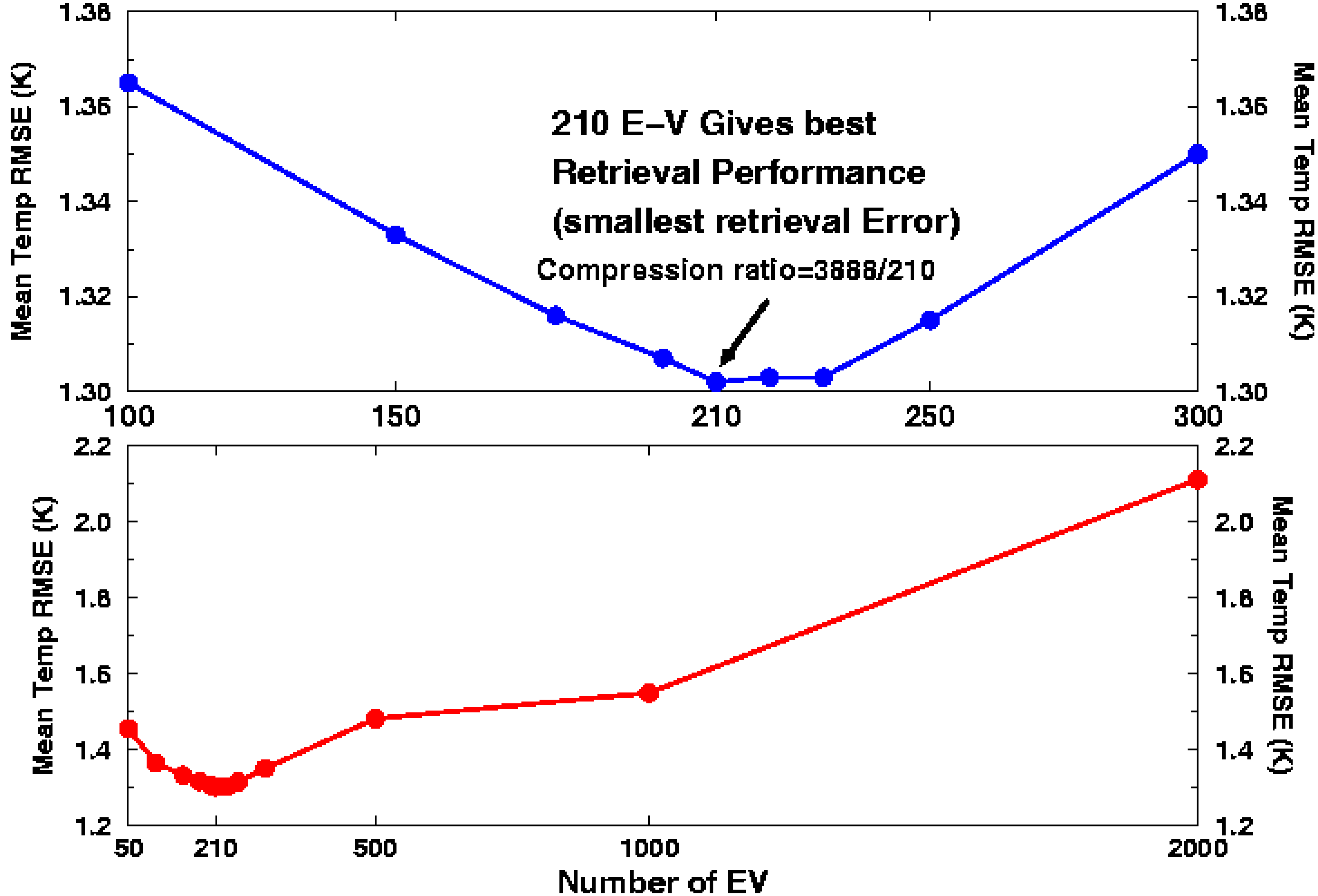
GIFTS Compression/Retrieval Analysis



C I M S S

GIFTS Data Compression Retrieval Error

Only Optimal Compression Can Improve Retrieval Performance



HPC On-board Compression Processing Requirement

- Computational Requirements
  - PC Generation
    - Mean : 2008*(41+1)=84336 flops=0.08 Mflops
    - COV : for 150*41*2 & 448*41 matrix -> 1.88*2+16.7 Mflops = 20.5 Mflops
    - SVD : 9.4*2+244 Mflops -> 262.8 Mflops
  - PC Compression
    - [630*(20+1)+150*(20+1)*2+448*(20+1)]*16384 -> 474 Mflops
  - Total = 0.08+20.5+262.8+474 = 757 Mflops

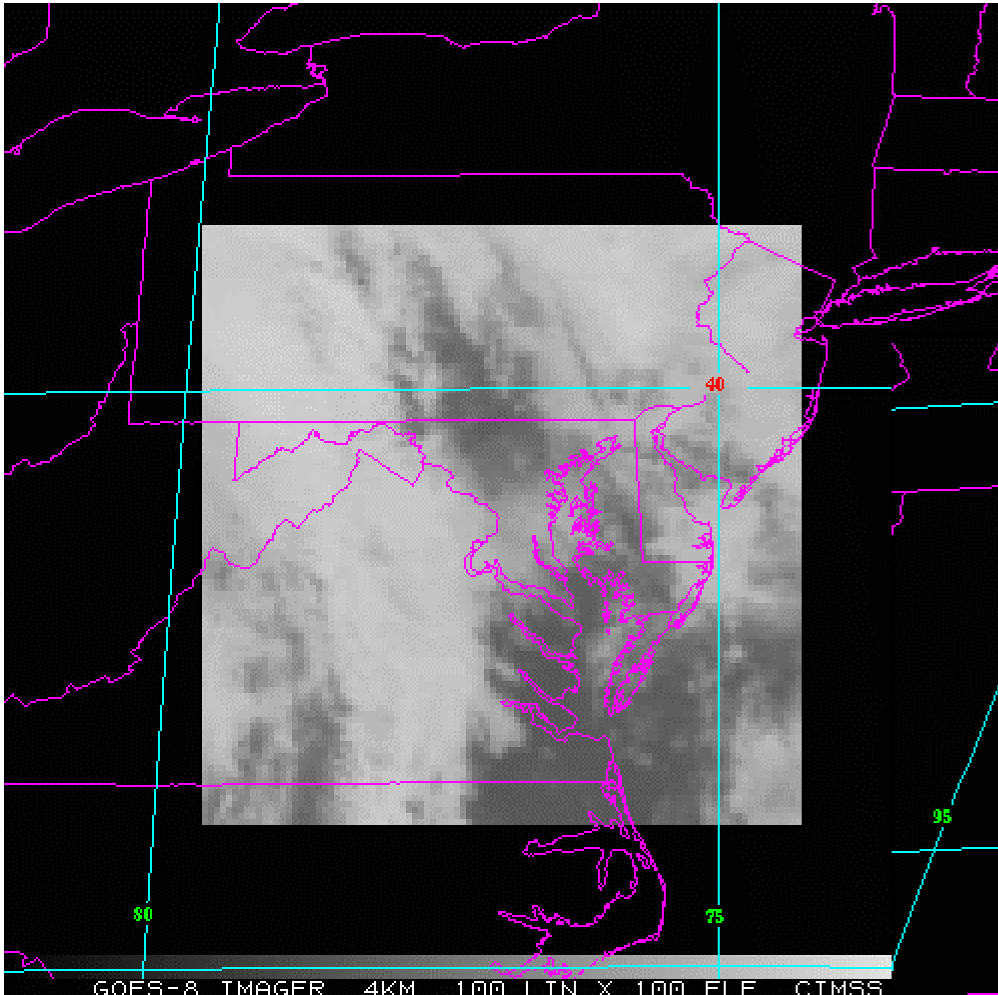
HPC On-Board Compression Processing Requirement

- Memory Requirements
  - PC Coefficients
    - Mean : 2008*16 = 0.03 Mbytes
    - COV : (630*630*16*2)+(150*150*16*2)+(448*448*16) = 16.62 Mbytes
    - PCs : 2*(20*630*16)+2*(20*150*16)+(20*448*16) = 0.64 Mbytes
  - Interferograms
    - Training : 41*2048*32 = 2.7 Mbytes
    - Cube data : 16384*2048*32 = 1073 Mbytes
  - Total = 1093 Mbytes

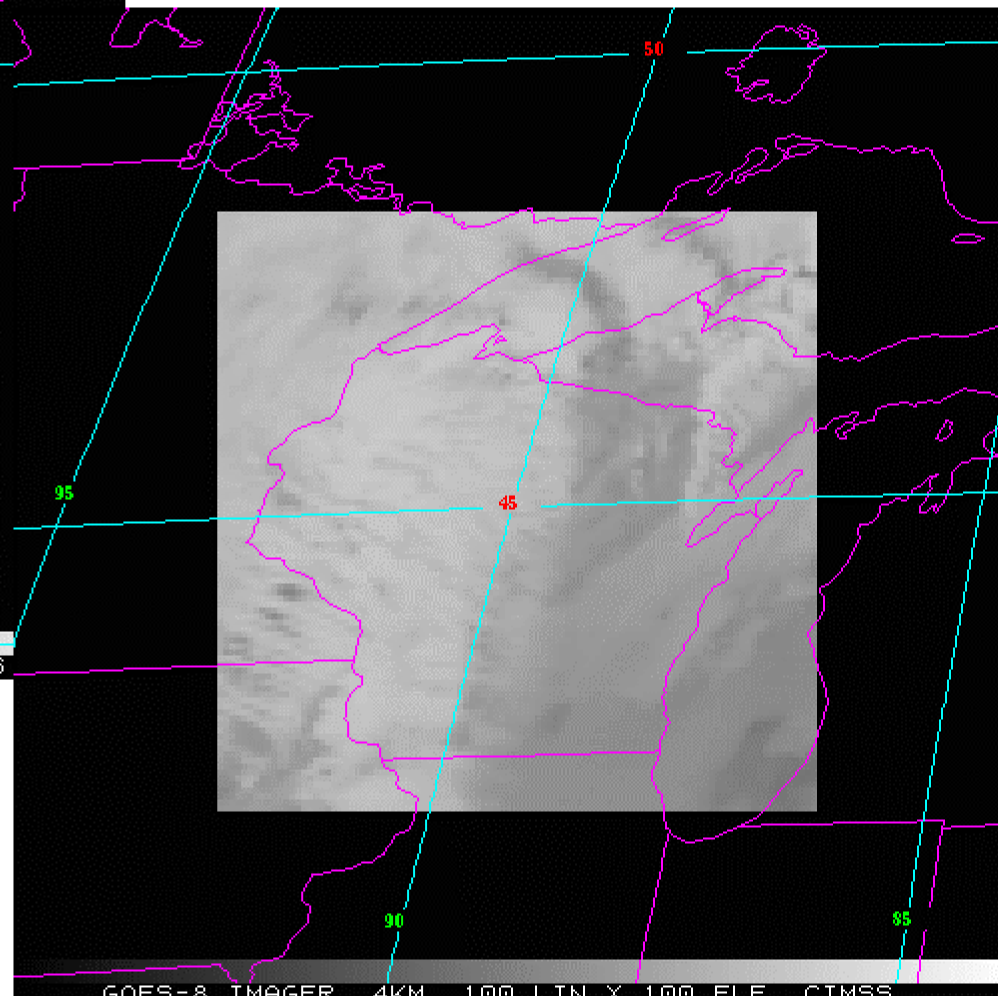
Onboard Computing Requirements

- Calibration
  - $4N + N\log_2(N)$ order calculation per interferogram
  - 15 ops per point LW; 16 ops per point S/MW
- Resampling from self-apodized state
  - $59N + 9N\log_2(N)$ order calculation per interferogram
  - 158 ops per point LW; 167 ops per point S/MW
  - can be mitigated by pre-interpolating raw interferograms before filtering
- Compression by Principal Component Analysis
  - $M*N$ order calculation per interferogram for M PCs
  - M=100 implies 100 ops per point

Potential DPC Processing Domain



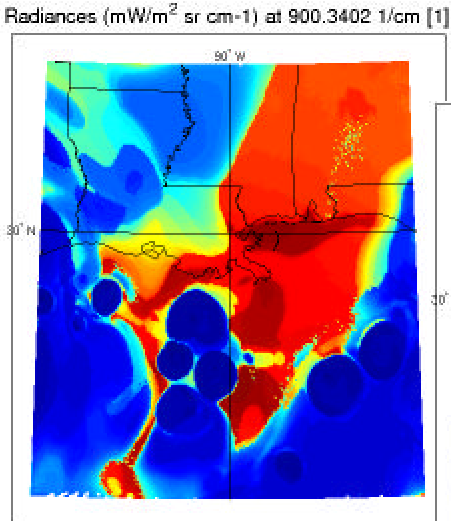
DC



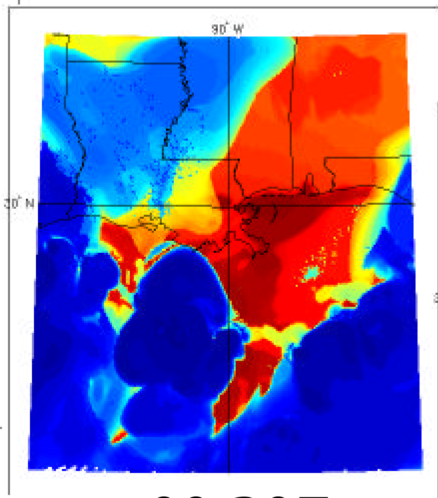
Madison

CIMSS

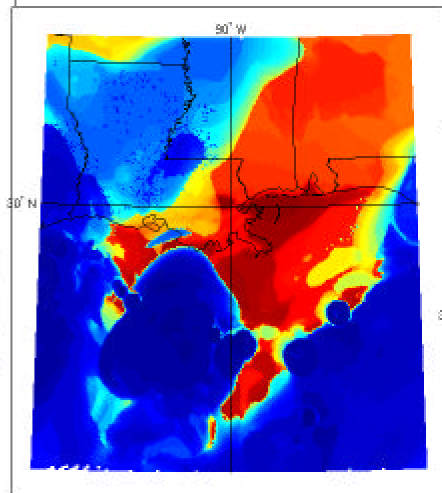
Potential ABS Dataset for Compression Study



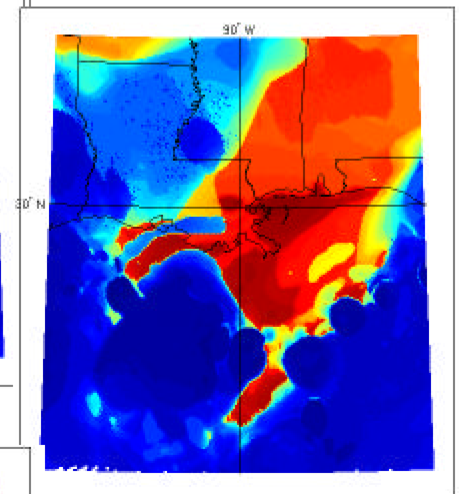
00:00Z



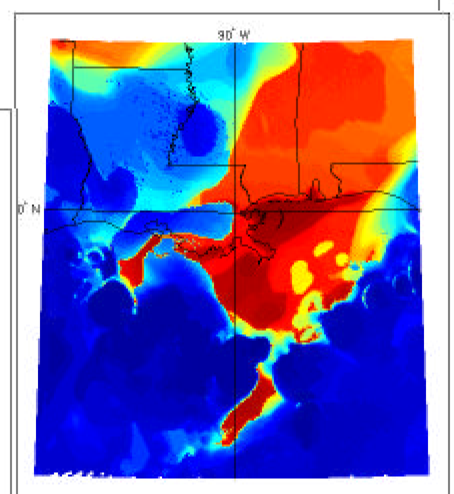
00:30Z



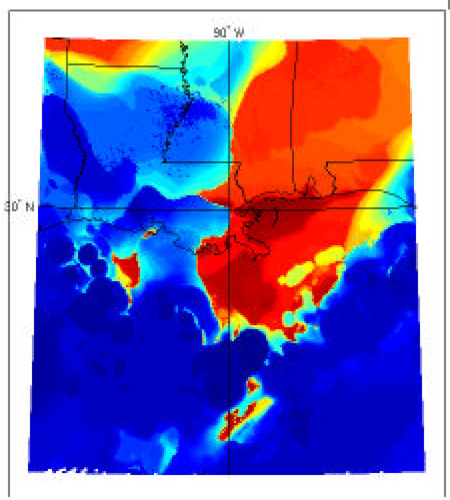
01:00Z



01:30Z



02:00Z



2:30Z

CIMSS

Implementation Roadmap for ABS Hyperspectral Data Compressing

Measurement Simulation:
Atmospheric Condition Modeling
TOA Radiance Modeling
Instrument Effect Modeling
Calibration Modeling

Processing Implementation:
On-Board
Ground-Based
Validation

Data Compression Study:
Raw/Calibrated Spectral Domain
Raw/Calibrated Spatial Domain
Dependent/Independent Domain
PC/EOF/SVD Transformation Approach
Wavelets Transformation Approach
Quantization/Coding
On-board/Ground-Based

Noise Estimates

CIMSS

Responsibility Roadmap for ABS Data Compressing

NOAA/Government:
Defined User Needs
Set Requirement
Coordinate Industry & Research Center

Industry:
Instrument Expertise
Space Qualified Parts
Design/Implementation

Research Centers:
Conduct Simulation
Conduct Processing Study
Define Processing Needs
Define Feasibility
Recommend Solution

CIMSS

Roadmap for ABS Data Compressing Study

1. Leveraging of NASA GIFTS on-board compression work

1-1. Processing Algorithm - Continue to optimize DPCC approach

1-2. Processing Requirements - Refine requirement for ABS efficient ground-based and on-board compression

1-3. Measurement/Products Impacts - Reconfirm compression impacts on both level-1 and level-2 data

CIMSS

Roadmap for ABS Data Compressing Study - continue

2. Adapting GIFTS Compression work with Enhancement/Improvement

2-1. Test enhancement of numerical method of deriving DPCC used in GIFTS experiment

2-2. Test variety of similar (non-PC) numerical techniques and inter-compare their respective performance and efficiency

3. Comparing Image and Spectral Compression Technique

3-1. Compare product impact of image (I.e. JPG 2000) and interferogram/spectral domain compression

CIMSS

Roadmap for ABS Data Compressing Study - continue

4. Modeling ABS Instrument Measurements

4-1. Develop ABS specific radiative transfer model

4-2. Model ABS instrument outputs - both signal & noise

4-3. Simulate ABS geo-orbital high temporal and spatially coherent measurements using high spatial resolution numerical model output and/or data cubes from field experiments

4-4. Modeling ABS pre-processing and calibration procedure

4-5. Provide ABS raw interferogram and calibrated spectral images

CIMSS

Roadmap for ABS Data Compressing Study - continue

5. Define Tradeoff Disadvantages/Benefits for "Lossless" and Lossy" Compression Approaches

6. Extended ABS compression study – use not only the data transformation based compression approach and to include all other compression components

6-1. Data quantization

6-2. Data coding and decoding

7. Define Performance for both Ground-based and On-board Compression

8. Make Recommendations for ABS Operational Consideration for Data Downlink and Ground Data Redistribution

CIMSS